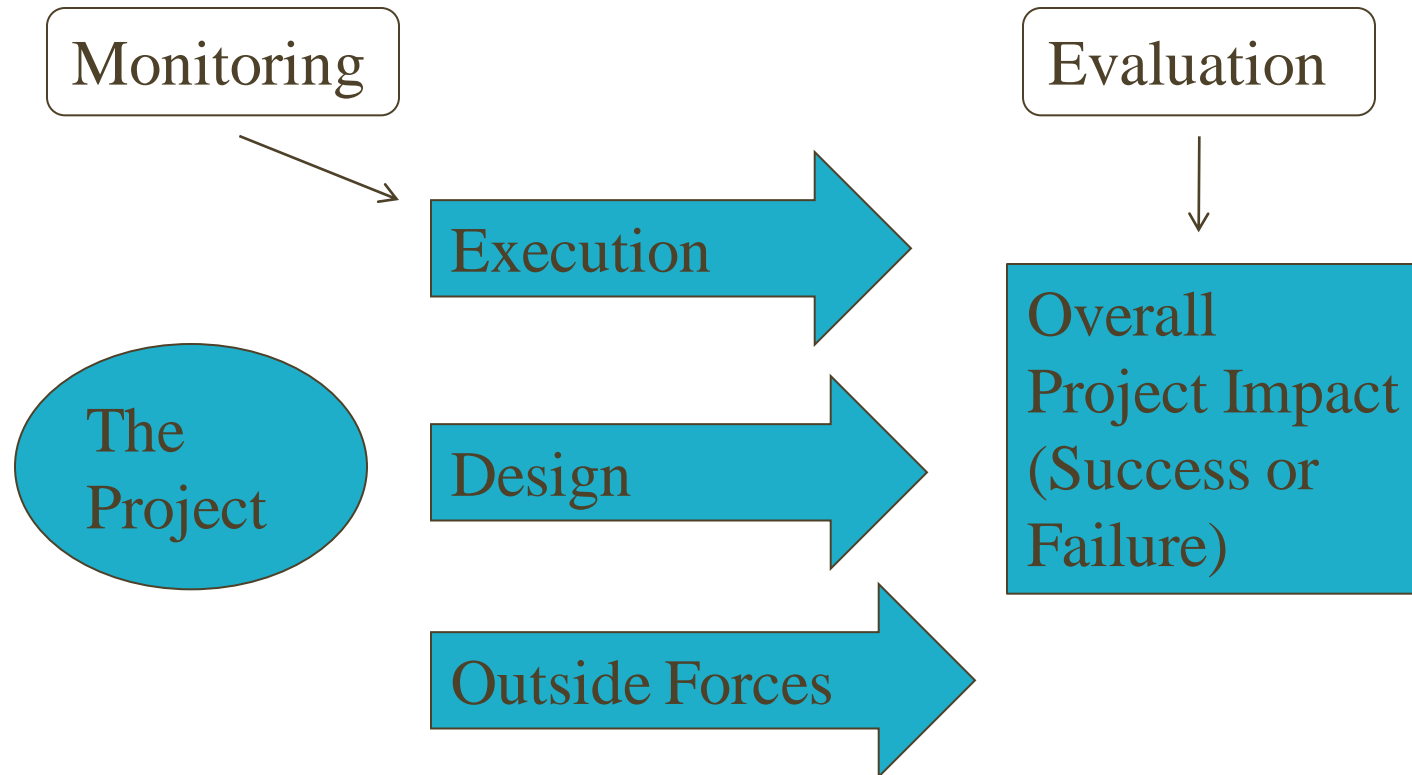# Agricultural Development Program: Lessons Learned from the Project Evaluation in Burkina Faso

Kenneth Leonard

University of Maryland

July 21st, 2011

Monitoring and Evaluation (M&E)

Monitoring

Evaluation

The Project

Execution

Design

Outside Forces

Overall Project Impact (Success or Failure)

Most people would rather be monitored than evaluated

- **Evaluation has more risk.**

- **Implementers have less control over outcomes than over execution and they like to be judged on the basis of what they can do**

- **Many projects are correctly implemented but fail whereas few are incorrectly implemented but succeed: Monitoring is a lower hurdle.**

First Lesson: M&E Implementers want M, Evaluators deliver E

- **The AD team in BF demanded surveys that measured the weight of cows and counted the number of vaccines distributed.**

- **Our team wanted incomes**

- **The Data Collection contractor answers to the AD team, not to the Evaluation team.**

- **The AD team thought we were there to monitor and did not understand why we were evaluating**

- **MCC cannot force MCA-BF to follow our recommendations.**

Evaluation is much more important than Monitoring

- The lessons of development are that many projects have not had the projected impacts despite excellent implementation

- We are in the development business, not the implementation business

- Design is crucial and understanding how to build projects that are not vulnerable to outside forces is also key

- Evaluation and the threat of evaluation brings a rigor to implementation and design that can improve the odds of success.

- Properly designed impact evaluation can improve implementation.

Evaluation is Good, but does it need to be Rigorous?

- **Rigorous Evaluation answers the following question, "if the program is rolled out or expanded, what impact can we expect?"**

  - Randomization across control and treatment is almost the only way to properly answer this question

- **What if the program is the best possible program to solve a specific development problem?**

  - An identical program will never be implemented anywhere else.

  - All potential beneficiaries in the study area will have received the benefits already.

  - The next thing to do is design the best project for another problem.

If it is the best program for specific problem …

- **That is a good thing; replicability is not the goal of development programs, development is the goal of development programs.**

- **This is not a medicine trial.**

- **No one cares what lessons can be learned and then used in other countries.**

  - Telling Burkina Faso that they should do something so that Mali can learn from their mistakes is insulting.

Conflict between designer and evaluator

- **Selection of beneficiaries:**

  - Designer has picked "best" regions, villages or individuals
  - Evaluator wants randomly selected regions, villages or individuals

- **Roll out of program:**

  - Designer wants optimized roll out for low cost and internal adaptation to the inevitable small problems
  - Evaluator wants randomized roll out

- **Implementation:**

  - Designer wants the best people and priority for the program
  - Evaluator wants something that is replicable (what the program might look like when excitement is gone)

- **Complexity:**

  - Designer wants the best possible combination of many independent interventions
  - Evaluator wants something very simple from which lessons can be learned.

Rigorous Evaluation may lead to the failure of the program

- Why implement in villages you know will fail?

- Why use extension officers you don't trust?

- Why restrict your ability to adapt the program to local conditions?

- Why make the team start in a random place rather than the best place?

- Why choose small incremental changes when you know the problem can only be solved by an integrated approach?

What does the Implementer want?

- **Do your "matching magic" and find a perfect counterfactual somewhere very far away from my project.**

- **Then compare the gains I show in my area to what we know must be stagnant development in the control area.**

- **What's wrong with this?**

  - If the treatment area was deliberately chosen, matching fails almost by definition; you know, a priori, that treatment and control are not the same.

  - It is not true that incomes are stagnant in the absence of a development program. Variability is the only constant of agricultural incomes.

Difficulties with Design

- **You are evaluating the implementer and, in fairness you must let him do his job**

- **There is no counterfactual or control group:**

    - All beneficiaries have been specifically chosen to maximize success

    - Matching assumes the implementer made a mistake and missed some good beneficiaries

    - Discontinuity matches the last people in with the first people excluded—this could be the group with the lowest possible impact.

- **One year impact is not meaningful**

    - Season to season comparisons are one year minimum

    - Farmers learn, expand, adapt

    - Excitement wears off

Tradeoff

- **"We had to destroy the program in order to evaluate it"**

VS.

- **We collected very little data in order to be sure our matching algorithms found sufficient common support and we pretended the people who chose villages didn't really know what they were doing**

Or is there something better?

Second Lesson: Evaluate the Implementer not just the program

- Learn about the design and selection process

- Find out what why he is proposing a particular order for implementation

- How does he plan to adapt the program as he goes along?

- What details of the implementation does his think will be instrumental in the success of the program?

- In BF, all decisions had been made by the time we arrived and the team that made the decisions had been replaced.

Third Lesson:
Insist on High Quality Data

- **Income, Yield, Profit measured correctly**

  - People may change crops, plots and labor allocation
  - Profit (income) is what matters, not production

- **Collect data for multiple years (not just before and after)**

  - You can follow a group for four years after receiving benefits

- **Collect data on the process of selection**

Fourth Lesson:
Use Randomization to Create Waves that can be Compared over time

- **Basically extended randomized roll out.**

- **Use the selection process to create groups of similar villages**

- **Randomly select which ones will receive benefits in which wave**

- **Evaluate the waves over time**

  - First wave might get four years of evaluation
  - Second wave might get only three

Fifth Lesson:
Do not allow for ad hoc updating of the program

- **No "mulligan's" for the first wave.**

- **If the implementer doesn't know how to do something,**

  - test it explicitly, randomizing across treatment and control.

  - Or use a pilot

- **Updating can be done if it is systematic and expected**

- **If the program improves significantly from wave to wave the evaluation cannot effectively measure this improvement.**

Multi Year Evaluation with Randomized Roll Out

- **Deliberate selection process: project designer chose farmers directly on the basis of potential success**

- **No counterfactual: all eligible farmers received the program**

- **Dynamic benefits: The program expected two year gains to be different from four year gains**

- **Profitability of resources devoted to the program was only part of the expected benefits**

- **Quasi-random selection of beneficiaries into waves**

- **Multi-year data**

Average Income
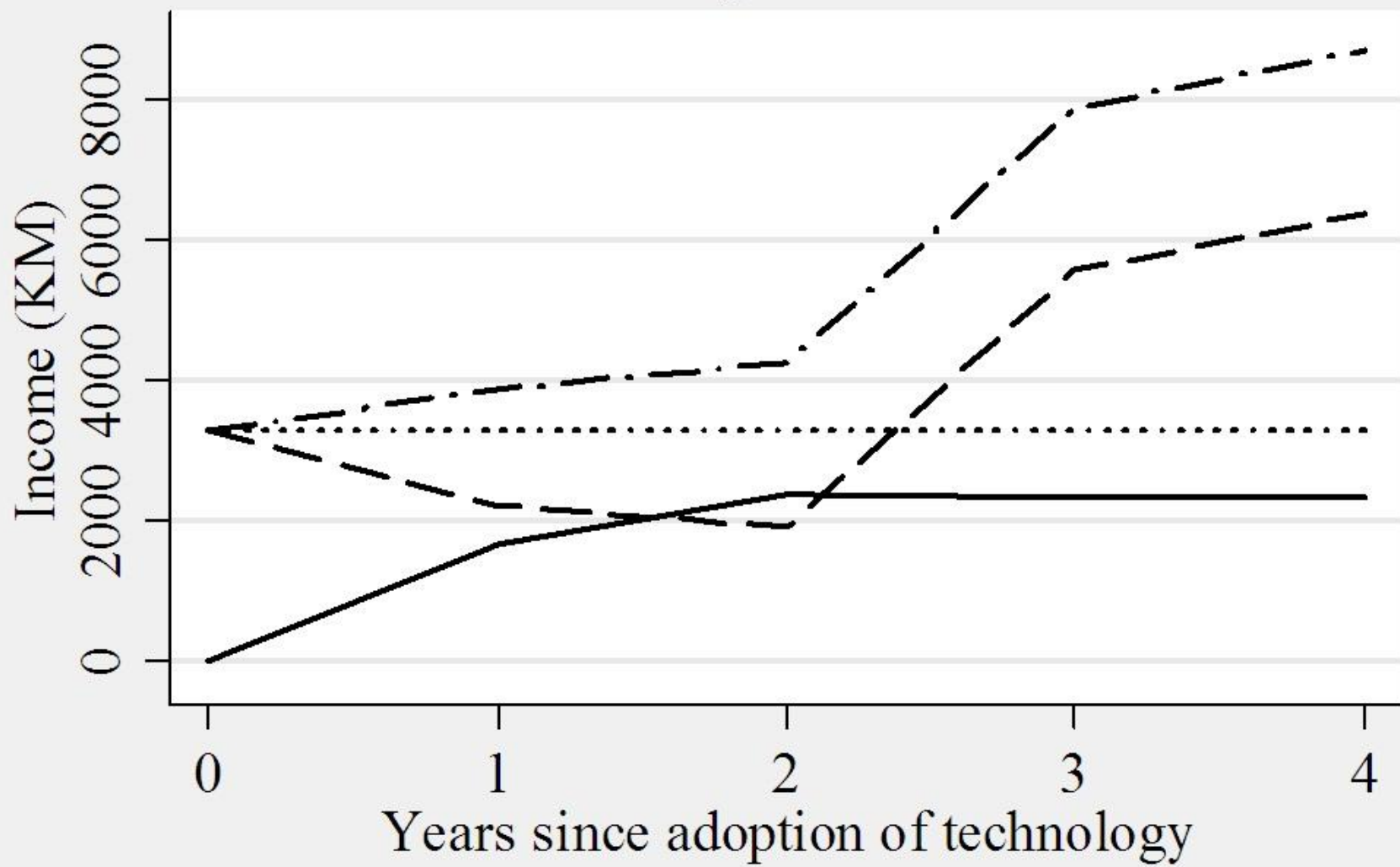
What this graph shows

- The previous data is not from BF, but from a program in Bosnia-Herzogovinia. Objectives of implementer were similar to BF, but were explicitly stated and modeled.

- **The program itself**

  - had a small but positive impact in the first year

  - It improved again in the second year

  - Total program benefits after 5 years do not exceed program cost

- **The program also affected other activities**

  - Income from other activities fell in the first year of the program

  - Income from other activities increased after the second year

  - Total gains in both the program and other activities exceeded the cost of the program.

What ended up happening in BF

- **No Randomized Rollout**

- **Over selection of potential control villages**

- **Significant data collection effort to explain the selection process and attempt to replicate it in control areas. (find villages that "would have been selected")**

- **Initial Matching on selection criteria across treatment and control areas**

- **Subsequent additional matching on the basis of experienced rainfall.**

- **Final double difference based on extensive matching**

- **If we can't match in a satisfactory manner, the evaluation will collapse.**

What was lost in the final plan?

- **We are left with an evaluation strategy that works only if we are lucky in matching or if we close our eyes to the real selection process. High risk or very poor design.**

- **Opportunity to find out why the program worked or did not work. All aspects are combined, no ability to follow gains over time.**

- **Opportunity to understand the impact that process (selection and management) can have on success.**